Research project report: Modelling MAPK/ERK pathway using Ordinary Differential Equations

Przemysław Pilipczuk

Abstract text

October 13, 2025

Preface text

Contents

1. Abstract

The MAPK/ERK pathway serves a key role in determining fate of a cell from extracellular inputs. Different cell fates are linked to a dynamics of the final node in the cascade, `ERK` kinases. To understand how different dynamc patterns of their behavior arise, we attempt to mechanistically model the simplified version of this pathway using Ordinary Differential Equations. Using data from a fibroblast cell transfected with optoRTK constructs and stimulated with various patterns of light, we attempted to estimated model parameters based on these experiments and use cross-validation techniques to check for generalization. Resulting fits had problems with achieving several performance metrics that would singify an optimal solution.

2. Introduction

TODO

3. Methods & Materials

3.1. Experimental Data

Data used for fitting the model was obtained from an experimental setup consisting of fibroblast cells transfected with an optoEGFR construct (optogenetic actuator) and being stimulated with various patterns of light. Data from 3 experiments were used in our pipeline, representing distinct light activation patterns. All experiments start with a 10 minute period of no actication, that we use for calibrating the baseline levels of activity.

First pattern that was investigated was a transient actuation pulse for a given amount of time. The experiment was repeated with different durations, namely $t = \{0, 50, 100, 200, 500, 1000\}$ ms. This experiment was used during development for testing and verifying due to its simplicity.
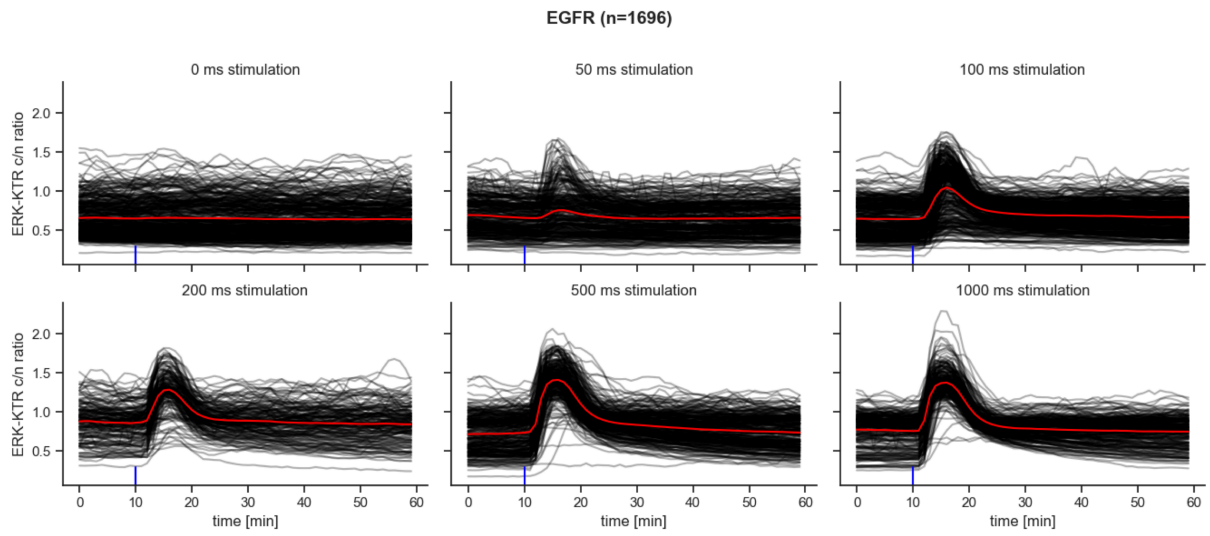

Figure 1: Single cell ERK-KTR activity in transient activation experiment. Each black curve represents a trajectory of a single cell. Red line is a median of all trajectories.

Second pattern that was incorporated was sustained activation. In this experiment, stimulation was started and sustained for a 140 minutes with three different power settings.
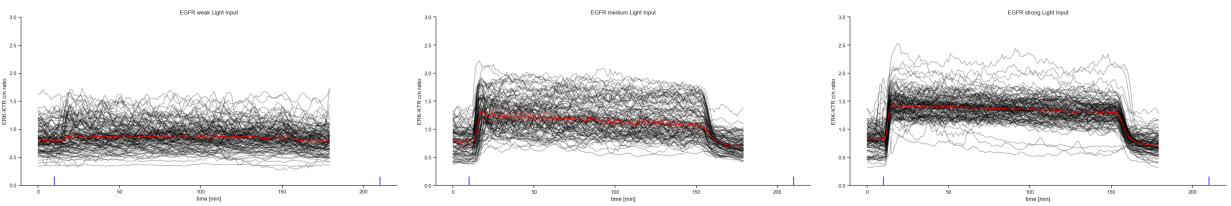

Figure 2: Single cell ERK-KTR activity in sustained activation experiment. Each black curve is a single cell trajectory of ERK-KTR, red line describes median.

Third pattern incorporated into the pipeline was a ramp pattern. This pattern starts by running a light pulse every minute, and increasing the pulse widht with every subsequent actication, starting from 0 and ending at 700ms pulse width, resulting in overall ramp shape of the activation curve over time.
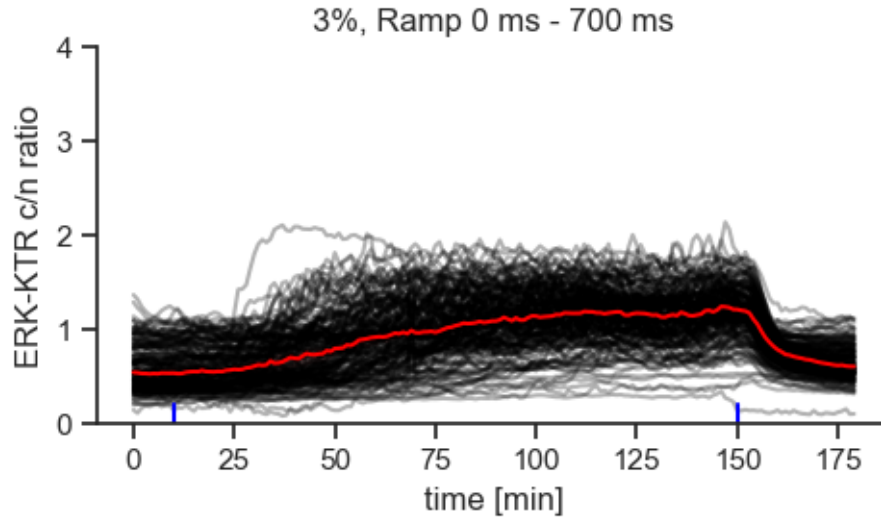
Figure 3: Single cell ERK-KTR activity in sustained activation experiment.
Each black curve is a single cell trajectory of ERK-KTR, red line describes median.

The development and initial testing of modeling pipeline was done using only transient activation experiment. Sustained and Ramp were incorporated as a means of providing cross-validation for the fitting process.

3.2. The chosen model

Finding the balance between simplicity and mechanistic accuracy is the key problem in model definition. Too simple model results in non-relevant findings, since it fails to capture complexity of the problem. Too complex model is harder to operate, and is prone to fragile to comlexity explosions, parameter unidentifiability, and poor choice of starting condition leading to nonsensical results.
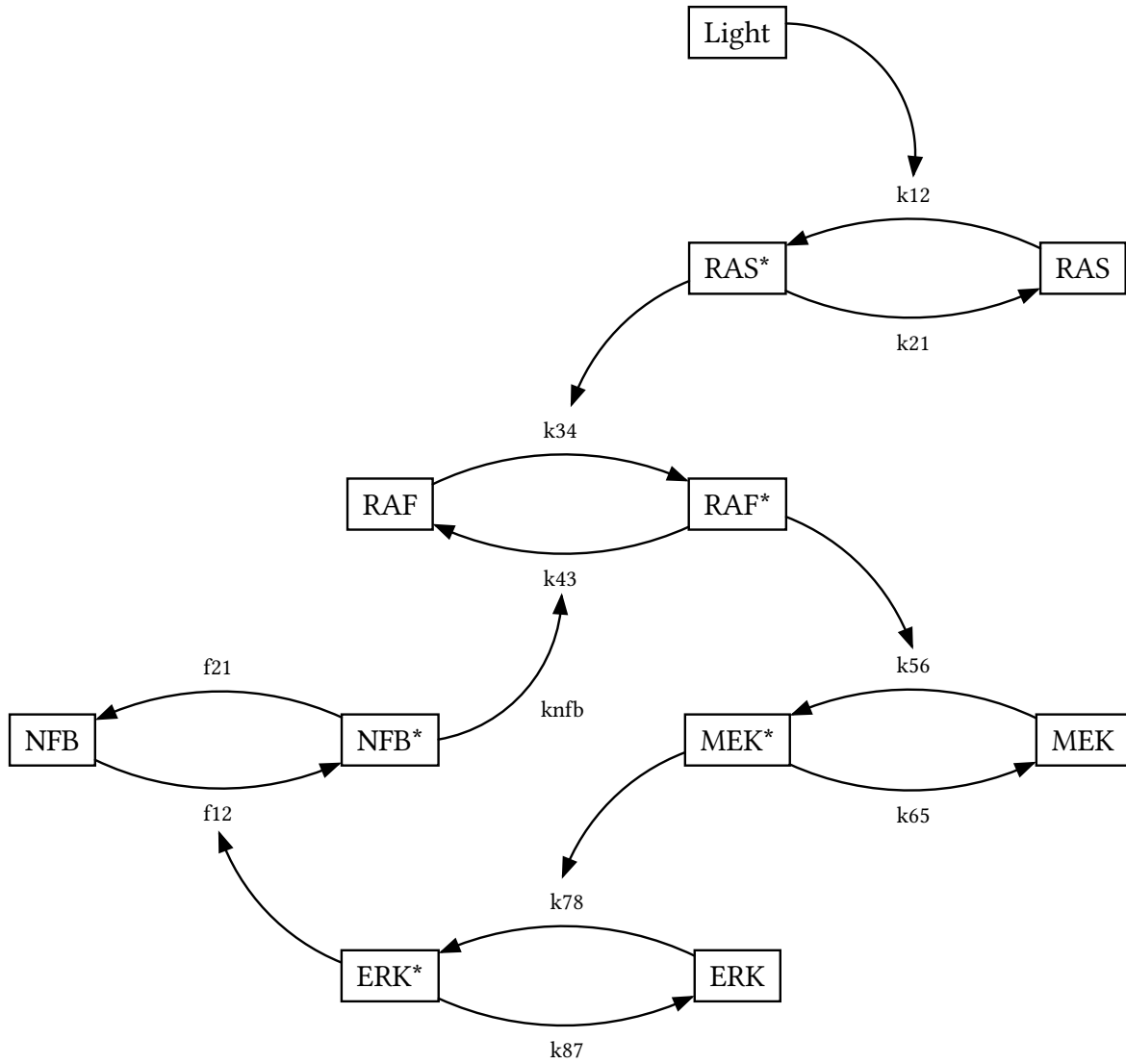
Figure 4: Diagram of the used model, representing simplified MAPK/ERK cascade.

Our chosen model starts at the level of RAS, and does not include the receptor layer. It treats light as a positive term in converting RAS to RAS*. The standard MAPK/ERK cascade of RAS-RAF-MEK-ERK is preserved We group whole families of kinases together and represent them as one vertex to preserve ontological grouping while minimising complexity of the model. Model also contains a negative feelback loop represented as a seperate NFB vertex in above graph. Modeling negative feedback as a separate state variable is not mechanistically accurate, but allows for a number of useful properties. It helps with model understandability, as negative feedback has clear and separate parameters that can be tweaked in isolation, while its existance as a state variable in the simulated model helps to track the contribution of negative feedback on the rest of the system. Another remarkable simplification of experimental setup comes from removal of reporter later (KTR-ERK complex). Presented model assumes that gathered experimental data coming from the reporter is a perfect proxy of actual state of ERK concentration in the cell.

3.3. Ordinary Differential Equations

Ordinary differential equations is a mathemathical framework for describing the change of one variable (dependant variable, here Y) over another variable (independent variable, here

X). Its simplest formulation is an expression that equates some mathematical expression to a derivative of our variable in question Y over the independent variable X.

$$\frac{dY}{dX} = \ldots \tag{1}$$

A solution to a differential equation is usually understood as obtaining a $Y(X)$ form, also called an *general solution*. General solutions can be obtained using an analytical solving process, the difficulty of which is heavily dependant on the specifc problem being solved. An often encountered problem with analytical solution approach is a problem that contains complex, nonlinearly coupled equations, which do not yield easily to this method. An alternative approach to solving a system that has these characteristics is a numerical one. This method relies on a provided initial conditions and a `X-step` resolution to simulate a single trajectory within an ODE system by evaluating the equations sequentially at a consecutive `X-steps` away from the provided starting point. This method, while providing a weaker form of solution, can deal with harder problems, including those that describe complex, nonlinear systems.

Basic formulation of our model as a set of differential equations comes down to expressing each family of kinases as a state variable and expressing interactions between these kinases as a positive or negative terms in their respective differential equations.

$$\frac{dRAS^*}{dt} = \text{light} * \left(\frac{RAS}{K\_12 + RAS} \right) - k\_21 * \left(\frac{RAS^*}{K\_21 + RAS^*} \right)$$

$$\frac{dRAF^*}{dt} = k\_34 * RAS^* * \left(\frac{RAF}{K\_34 + RAF} \right) - (knfb * NFB^* + k\_43) * \left(\frac{RAF^*}{K\_43 + RAF^*} \right)$$

$$\frac{dMEK^*}{dt} = k\_56 * RAF^* * \left(\frac{MEK}{K\_56 + MEK} \right) - k\_65 * \left(\frac{MEK^*}{K\_65 + MEK^*} \right) \tag{2}$$

$$\frac{dNFB^*}{dt} = f\_12 * ERK^* * \left(\frac{NFB}{F\_12 + NFB} \right) - f\_21 * \left(\frac{NFB^*}{F\_21 + NFB^*} \right)$$

$$\frac{dERK^*}{dt} = k\_78 * MEK^* * \left(\frac{ERK}{K\_78 + ERK} \right) - k\_87 * \left(\frac{ERK^*}{K\_87 + ERK^*} \right)$$

3.4. Modeling & Simulation pipeline

Model definition was done in symbolic form using SymPy in python (TODO: reference them), as a list of symbolic differential equations. The same tool was also used to lower the symbolic expression down to a numerical representation (`lambdify` function). A constructed model consists of a set of differential equations and parameter values. An effort was made to make this pipeline model-and-experiment agnostic and have low friction for implementing new models. This effort led to an architecture where each new model needs only to fulfill a simple interface for defining its equations and parameters, and each new experiment needs to define its own pattern of input and a parsing function for the data it generated.

An interactive environment for simulating behavior and exploring perturbations to parameters was constructed using marimo notebooks.

Parameter estimation was done using tools from SciPy library, namely `minimize` and `solve_ivp`

to optimize loss function and to solve a numerical system respectively. The general scheme involved minimizing residual sum of squares for the data given in an experiment across all the groups in a given experiment.

4. Results

5. Discussion

5.1. Time resolution trouble

During implementation of the parameter estimation regime, a technical difficulty was encountered. The experiments had been replicated multiple times with different parameter values for light stimulation, usually differing in stimulation time of a pulse of light, with values ranging from 50ms to 2000ms. At the same time, data for those experiments was collected with a rate of one sample per minute. Such difference in time ranges between sampling and input means that our experimental data will never be able to differentiate between the light stimulation duration groups within an experiment just by looking at the data (we would need to increase the sampling rate 1200 times to be able to discern signals that lasted 50ms, the most fine-grained light signal in our experimental data). To circumvent this problem, two solutions were contemplated.

1. An interpolation algorithm could have been used to synthetically create data points of the sampling rate we desire. This forces our new data points to inherit our assumptions about how the data is interpolated, which could introduce a systematic bias (as example, for linearity).

2. A proxy metric for a total light energy of a light pulse for a datapoint is introduced instead of a "binary" baseline. This allows for more flexibility but at a cost of having to create a realistically-scaling function that works for all kinds of experiments and across different levels of magnitude of light stimulation duration.

After theoretical evaluation of pros and const of both, a second solution strategy was picked, due to a better compatibility with the goal of having short training times.

6. Conclusions

7. Future work

8. Supplementary material

Bibliography

Index of Figures